



# **DSMM Use Case Studies**

## **– Briefing to ESIP Data Stewardship Committee**

**Ge Peng, Ph.D.**

Cooperative Institute for Climate and Satellites – North Carolina C (CICS-NC),  
North Carolina State University and NOAA's  
National Centers for Environmental Information (NCEI)

**18 April 2016**





# What Is the NCEI/CICS-NC Scientific Data Stewardship Maturity Matrix (DSMM)?

*A Unified Framework for  
Measuring **Stewardship** Practices Applied to  
Individual Digital Earth Sciences Data Products*

Developed Jointly  
by Data Scientists and Subject Matter Experts from CICS-NC and NCEI,  
Leveraging Institutional Knowledge and Community Best Practices and Standards



# DSMM Defines Measureable, Five-Level Progressive Practices in *Nine* Quasi-Independent Key Components

<div>Maturity Scale</div> <div>Key Component</div>	Level 1 - Ad Hoc Not Managed	Level 2 - Minimal Managed Limited	Level 3 - Intermediate Managed Defined, Partially Implemented	Level 4 - Advanced Managed Well-Defined, Fully Implemented	Level 5 - Optimal Level 4 + Measured , Controlled , Audit
Preservability	<i>The state of being preservable</i>				
Accessibility	<i>The state of being publicly searchable and accessible</i>				
Usability	<i>The state of data product being easy to understand and use</i>				
Production Sustainability	<i>The state of data production being sustainable and extendable</i>				
Data Quality Assurance	<i>The state of data product quality being assured/screened</i>				
Data Quality Control /Monitoring	<i>The state of data product quality being controlled and monitored</i>				
Data Quality Assessment	<i>The state of data product quality being assessed</i>				
Transparency /Traceability	<i>The state of being transparent, trackable, and traceable</i>				
Data Integrity	<i>The state of data integrity being verifiable</i>				


(more info: [www.tinyurl.com/DSMMintro](http://www.tinyurl.com/DSMMintro))



# ESIP DSMM Use Case Study

(Data Stewardship Committee/Information Quality Cluster)

## The Purposes

- Evaluate and improve DSMM and generalize its across-disciplines application
  - Evaluate and improve the DSMM self-assessment template
  - Develop community-wide dataset-centric stewardship requirements and guidance
- 



# Selected Datasets

Data Type	Dataset	Status
Model Reanalysis	NCAR Global Climate Four-Dimensional Data Assimilation Hourly 40km Reanalysis	Baselined *
Ecological Data	DataOne Member Node SBC LTER (Long Term Ecological Research) Network	Revised assessment draft review
Long-tail Data	NSF ACADIS (Advanced Cooperative Arctic Data and Information Service)	Initial assessment draft
Socioeconomic Data	NASA Socioeconomic Data	Initial assessment draft
Paleo Data	Borehole	Not yet assessed


\* The assessment status levels progress in the order of: Not yet assessed; Preliminary assessment; Initial assessment draft; Revised assessment draft; Final assessment draft; and Baselined.



# **NCEI Pilot DSMM Use Case Study**

**(NCEI Data Stewardship Division; Climate Science Division;  
NOAA Climate Data Record Program)**

## **The Purposes**

- Demonstrate the utility of DSMM, ensure and improve consistency of its application to various data types
  - Establish baselines for high-utility/impact NCEI core datasets (2-5 datasets)
  - Explore requirements for collecting and capturing content-rich quality metadata
- 



# Selected NCEI Core Datasets

Data Type	Dataset	Status
Satellite – polar ocean	NOAA/NSIDC Sea Ice Concentration CDR	Baselined
GIS - regional	NCEI-CO Digital Elevation Models (DEM)	Revised assessment draft
Station - in situ - land	GHCN-M	Baselined
Station - gridded - land	National Climate Division (nCliDiv)	Not yet started
Satellite – global ocean	OI SST CDR	Final assessment draft review
Physical Records - In Situ Monthly Summaries	Local Climatological Data	Initial assessment draft review
Paleo – global land	International Tree-Ring Data Bank (ITRDB)	Final assessment draft review



# NOAA Application of DSMM

## (NOAA *OneStop* Program)

**The Purpose of *OneStop* – Improve the discovery, access, and usability of NOAA Data**

- Help improve NOAA data discoverability and usability,
  - provide ranking information for search and discovery,
  - provide content-rich quality information for data usability.





# Thank You

- Everyone who has led, coordinated, or participated in the ESIP & NCEI use case studies

## Quick Links to DSMM Related Resources:

- **ESIP DSMM use case wiki page:**
    - [wiki.esipfed.org/index.php/Data\\_Maturity\\_Matrix](http://wiki.esipfed.org/index.php/Data_Maturity_Matrix)
  - **Self-Assessment Template:**
    - [www.tinyurl.com/DSMMtemplate](http://www.tinyurl.com/DSMMtemplate)
    - [https://figshare.com/articles/NCDC\\_CICSNC\\_SDSMM\\_Template/1211954](https://figshare.com/articles/NCDC_CICSNC_SDSMM_Template/1211954)
  - **High-level background slides & DSMM in pdf:**
    - [www.tinyurl.com/DSMMintro](http://www.tinyurl.com/DSMMintro)
    - [www.tinyurl.com/DSMMslide](http://www.tinyurl.com/DSMMslide)
- 



## Sessions of interests at the ESIP 2016 Summer Meeting:

- **<http://commons.esipfed.org/node/9036>**

Presentations and panel discussion on lifecycle stages-based reference frameworks for assessing maturity of individual digital Earth Science data products

- **<http://commons.esipfed.org/node/9035>**

Maturity models, not just DSMM, use case study update and working/training





## Categories of Maturity Assessments Utilizing a Maturity Model

- **Category 1:** No assessment done.
- **Category 2:** Self-assessment – preliminary assessment carried out by individual for internal or personal use; results (ratings and justifications) are not publicly available; abiding to non-disclosure agreement.
- **Category 3:** Internal evaluation – complete assessment carried out by a person, group, or institution, and reviewed internally; results (ratings and justifications) are publicly available.
- **Category 4:** Independent evaluation – complete evaluation carried out and/or reviewed by an independent person, group, or institution; The level of expertise of the evaluator in the maturity model utilized is described; results (ratings and justifications) are publicly available; Improvement recommendations may be provided.
- **Category 5:** Certified evaluation – complete evaluation reviewed and certified by an established authoritative person, group, or institution; results (ratings and justifications) and improvement recommendations are publicly available; Maturity update frequency is defined and implemented.




# **Back Up Slides**






# Main Benefits and Lessons Learned

- **It is useful to have a reference framework**
    - Actionable practices defined in the DSMM help & make it easier to define stewardship maturity requirements,
  - **Assessment is useful in identifying the strengths and weakness of both the dataset and organization capability**
    - Uncover potential gaps and ensure across-domain integration, e.g., data product web site, DOI landing page, data application site.
  - **Simple does not mean easy**
    - Strong dependency on experience, domain knowledge and availability of information.
  - **Well-defined and understood data management and stewardship processes and standardized document(s) associated with the processes help greatly reduce the required effort**
- 



# Challenges

- **Information** on practices applied to a dataset during different life stages of the data set **is needed** but often **not collected or documented**, or **documented in a way** that can be easily found or understood by people outside of its domain/field.
  - **Ensuring and improving** data stewardship maturity requires a **coordinated, across-multiple-domains** effort.
  - **Long-tail data**
    - Variety in practices, often not documented,
    - Static,
    - Assessment for individual datasets in the collection or the derived product?
- 

# The Scope of Stewardship Practices

- Those applied to individual datasets – measureable and progressive
- Those associated with the functional entities of the Open Archival Information System (OAIS) (within the shaded box in the diagram below)

